

Improving Durbin's method to estimate MA processes

Maximilian Ludwig*

this version: April 12, 2013

Abstract

This paper provides a simple method to estimate both univariate and multivariate MA processes. Similar to Durbin's method, it rests on the recursive relation between the parameters of the MA process and those of its AR representation. This recursive relation is shown to be valid both for invertible / stable and non invertible / unstable processes under the assumption that the process has no constant and started from zero. This makes the method suitable for unit root processes too.

1 Introduction and Summary

A classic method to estimate the parameters of pure MA processes is provided by Durbin (1959). This method is a two step estimation. First, an AR process of large order is estimated using OLS. The parameter estimates are then used to estimate the parameters of the MA process using the Yule-Walker estimator, which is shown to be efficient for this purpose by Durbin.

In this paper, I show that under one assumption, the well know recursive formulae to compute the MA representation of an AR process / the AR representation of an MA process can be derived without assuming stability of the AR / invertibility of the MA process. Inspecting the recursive formula reveals that the AR parameters can be viewed as generated by an AR process parameterized by the MA. Therefore Durbin's method is basically a Yule-Walker estimation of the latter $AR(q)$ process on parameters from the initial AR estimation. It is however well known and documented by e.g. Sandgren and Stoica (2006) that the precision of estimates obtained by Durbin's method is unsatisfactory if one or more roots of the data generating MA process are close to the unit circle. The reason for this is simple: As documented by Tjostheim and Paulsen (1983) and others, the Yule-Walker estimator is not well suited to estimate AR processes with roots close to the unit circle and biased in case one root is on the unit circle. Since the OLS estimator is consistent also for

*University of Hamburg, Department of Economics, mludwig@mludwig.org. The latest version of this paper and a ready to use MatLab implementation of the estimator are available at <http://mludwig.org/research.html>.

AR processes with unit root, a natural modification of Durbin's method is the use of the OLS estimator for the second estimation. But since the Yule-Walker estimator is efficient, one faces a trade-off between a lower efficiency for MA processes with roots away from the unit circle and more reliable estimates in case of one or more roots close to unity. However, by exploiting the structural information implied by the recursive formula one can use a restricted OLS estimator to counteract this disadvantage. I perform a simulation for the parameter space of an MA(2) process which indicates that this approach yields substantial gains in efficiency in comparison to Durbin's method for a very large part of the investigated parameter space. Finally, a generalization of the approach for multivariate MA processes is provided.

2 The recursive formulae

2.1 A more general MA representation

The most important tool throughout the analysis is the so called companion form of an AR(p) process. This means that the process is noted as VAR(1):

$$\underbrace{\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix}}_{y_t} = \underbrace{\begin{bmatrix} v \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_v + \underbrace{\begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & 1 & 0 \end{bmatrix}}_F \underbrace{\begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix}}_{y_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\varepsilon_t} \quad (1)$$

For a given vector of starting values $y_{t-\tau}$ it is always possible to write the process as a function of errors terms, starting values and time:

$$\begin{aligned} y_{t-\tau+1} &= v + Fy_{t-\tau} + \varepsilon_{t-\tau+1} \\ y_{t-\tau+2} &= v + Fv + FFy_{t-\tau} + F\varepsilon_{t-\tau+1} + \varepsilon_{t-\tau+2} \\ y_{t-\tau+3} &= v + Fv + F^2v + FF^2y_{t-\tau} + F^2\varepsilon_{t-\tau+1} + F\varepsilon_{t-\tau+2} + \varepsilon_{t-\tau+3} \\ &\vdots \\ y_t &= \left(I + \sum_{j=1}^{\tau-1} F^j \right) v + F^\tau y_{t-\tau} + \sum_{j=1}^{\tau-1} F^j \varepsilon_{t-j} + \varepsilon_t \end{aligned}$$

Defining $F^0 \equiv I$ and using $(I - F)^{-1}(I - F)(\sum_{j=0}^{\tau-1} F^j) = (I - F)^{-1}(I - F^\tau)$ one can write this equation in a manner that I call the *generalized moving average representation*:

$$y_t = \underbrace{(I - F)^{-1}(I - F^\tau)v + F^\tau y_{t-\tau}}_{w_t} + \sum_{j=0}^{\tau-1} F^j \varepsilon_{t-j} \quad (2)$$

This name seems appropriate since the generalized moving average representation contains the textbook moving average representation in its companion form version as a special case. This is easy to see in case of distinct eigenvalues of F . It is well known that in this case, F can be decomposed as $W\Lambda W^{-1}$

where Λ is a diagonal matrix of the eigenvalues and W a matrix containing the eigenvectors. This allows to write $F^\tau = W\Lambda^\tau W^{-1}$. As shown by Hamilton (1994, pp.21) the eigenvalues of F are equivalent to the inverse roots of the lag polynomial and hence the process is stable if the norm of all eigenvalues of F is smaller than one. Then for $\tau \rightarrow \infty$ $\Lambda^\tau \rightarrow O_p$ where O_p denotes a matrix of zeros in $\mathbb{R}^{p \times p}$. Thus: $\lim_{\tau \rightarrow \infty} E[w_t] = (I - F)^{-1}v = \mu$, which allows to write (2) for $\tau = \infty$ and eigenvalues in the unit circle as:

$$y_t = \mu + \sum_{j=0}^{\infty} F^j \varepsilon_{t-j} \quad (3)$$

The literature rarely approaches AR processes with the companion form but uses lag polynomials. However, their use in case one seeks to invert a lag polynomial is linked to the assumption of $\tau = \infty$ and stability / invertibility. For the important special of a process with $v = 0$, these assumptions can be replaced with a single assumption that might be more sensible in some settings:

Assumption: *The "memory" of the process has been emptied at some time in the past, meaning $y_{t-\tau} = 0$.*

Under this assumption $w_t = 0$ for any parametrization and (2) looks quite similar to (3) with $\mu = 0$:

$$y_t = \sum_{j=0}^{\tau-1} F^j \varepsilon_{t-j} \quad (4)$$

However, (4) is an MA representation of both stable and unstable AR processes. From here on we shall impose both $v = 0$ and $y_{t-\tau} = 0$, where τ might be both finite or infinite.

2.2 The recursive formula for AR processes

To my best knowledge, the only way the recursive formula has been derived previously is by using the lag polynomial of the AR process, that is: assuming $\tau = \infty$ and stability. I shall reproduce this approach here for convenience. Consider an $AR(p)$ process without constant in lag notation: $y_t \phi(L) = \varepsilon_t$. Noting $\phi(L)^{-1} \equiv \psi(L)$ gives the identity:

$$\begin{aligned} 1 &= \phi(L)\psi(L) = (1 - \sum_{j=1}^p \phi_j L^j) (\sum_{i=0}^{\infty} \psi_i L^i) \\ &= (\sum_{i=0}^{\infty} \psi_i L^i) - \phi_1 L (\sum_{i=0}^{\infty} \psi_i L^i) - \dots - \phi_p L^p (\sum_{i=0}^{\infty} \psi_i L^i) \\ &= \psi_0 + (\psi_1 - \phi_1 \psi_0) L + (\psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0) L^2 + \dots \\ &\quad \dots + (\psi_p - \phi_1 \psi_{p-1} - \phi_2 \psi_{p-2} - \dots - \phi_p \psi_0) L^p \dots \\ &\quad \dots + (\psi_{p+1} - \phi_1 \psi_p - \dots - \phi_p \psi_1) L^{p+1} + \dots \end{aligned}$$

Since $\psi_0 \equiv 1$ the parameters of the MA representation can be recursively computed as $\psi_1 = \phi_1$, $\psi_2 = \phi_1 \psi_1 + \phi_2$ and so on.

To derive the same recursive formula from (4) one has to compute the powers of F , which seems harder than it actually is. Consider F^2 :

$$\begin{bmatrix} \phi_1 & \dots & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1 & \dots & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \phi_1^2 + \phi_2 & \dots & \phi_1 \phi_{p-2} + \phi_{p-1} & \phi_1 \phi_{p-1} + \phi_p & \phi_1 \phi_p \\ \phi_1 & \dots & \phi_{p-2} & \phi_{p-1} & \phi_p \\ 1 & \dots & 0 & 0 & 0 \\ 0 & \ddots & & 0 & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & & 1 & 0 & 0 \end{bmatrix} \quad (5)$$

Obviously, the first column contains the first parameters of the MA representation. Since all elements of ε_{t-j} except the first are zero, one can ignore the values in all columns of F^j except for the first one. Knowing F^2 and recalling that F^j is generated by the multiplication $F \times F^{j-1}$ allows to compute F^3 as:

$$F^3 = \begin{bmatrix} \phi_1 & \dots & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \psi_2 & \varpi & \dots & \varpi \\ \psi_1 & \varpi & \dots & \varpi \\ 1 & \varpi & \dots & \varpi \\ 0 & \varpi & \dots & \varpi \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \varpi & \dots & \varpi \end{bmatrix} = \begin{bmatrix} \phi_1 \psi_2 + \phi_2 \psi_1 + \phi_3 & \varpi & \dots & \varpi \\ \psi_2 & \varpi & \dots & \varpi \\ \psi_1 & \varpi & \dots & \varpi \\ 1 & \varpi & \dots & \varpi \\ 0 & \varpi & \dots & \varpi \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \varpi & \dots & \varpi \end{bmatrix}$$

where ϖ indicates values that are of no interest for the analysis. Thus, $\psi_3 = \phi_1 \psi_2 + \phi_2 \psi_1 + \phi_3$ and obviously one can compute F^4 and all F^j for $j = 2, \dots, \tau$ in a similar manner. Note that for $j > p$ one gets:

$$F^j = \begin{bmatrix} \phi_1 & \dots & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \psi_{j-1} & \varpi & \dots & \varpi \\ \psi_{j-2} & \varpi & \dots & \varpi \\ \vdots & \vdots & \vdots & \vdots \\ \psi_{j-p} & \varpi & \dots & \varpi \end{bmatrix}$$

Which means that for $j > p$, the MA parameters are generated by the difference equation $\psi_j = \sum_{i=1}^p \phi_i \psi_{j-i}$.

2.3 The recursive formula for MA processes

Using the Lag operator, a zero mean $MA(q)$ process can be noted as $y_t = \psi(L)\varepsilon_t$. The corresponding $AR(\infty)$ process $\phi(L)y_t = \varepsilon_t$ can be used in the same way as above to gain the identity

$$\begin{aligned} 1 &= \psi(L)\phi(L) = \phi(L) + \psi_1 L\phi(L) + \dots + \psi_q L^q \phi(L) \\ &= (1 - \sum_{j=1}^{\infty} \phi_j L^j) + \psi_1 L(1 - \sum_{j=1}^{\infty} \phi_j L^j) + \dots \\ \Leftrightarrow 0 &= - \sum_{j=1}^{\infty} \phi_j L^j + \psi_1 L - \sum_{j=1}^{\infty} \psi_1 \phi_j L^{j+1} + \psi_2 L^2 - \sum_{j=1}^{\infty} \psi_2 \phi_j L^{j+2} + \dots \\ \Leftrightarrow 0 &= (\psi_1 - \phi_1)L + (\psi_2 - \phi_2 - \psi_1 \phi_1)L^2 + \dots \\ &\quad \dots + (\psi_q - \phi_q - \psi_1 \phi_{q-1} - \psi_2 \phi_{q-2} - \dots - \psi_{q-1} \phi_1)L^q \dots \\ &\quad \dots + (-\phi_{q+1} - \psi_1 \phi_q - \dots - \psi_q \phi_1)L^{q+1} \dots \\ &\quad \dots + (-\phi_{q+2} - \psi_1 \phi_{q+1} - \dots - \psi_q \phi_2)L^{q+2} + \dots \end{aligned}$$

Thus, $\phi_1 = \psi_1$, $\psi_2 = \phi_2 + \psi_1 \phi_1$ and so on. To show that this is indeed the AR representation of both invertible and noninvertible MA processes given $v = 0$ and $y_{t-\tau} = 0$, one needs to note (4) compatible to the companion form of an

AR(p) process. Suppose we have $T > q$ observations of y_t , know that they were generated by an MA(q) process and look for an AR(l) representation of this process. To do so, write down the structure of the data generating process in a rather unusual way:

$$\begin{aligned}
\underbrace{\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-l} \end{bmatrix}}_{y_t} - \underbrace{\begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\varepsilon_t} &= \underbrace{\begin{bmatrix} \psi_1 & \varpi & \dots & \varpi \\ 1 & \varpi & \dots & \varpi \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \varpi & \dots & \varpi \end{bmatrix}}_{M_1} \begin{bmatrix} \varepsilon_{t-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \underbrace{\begin{bmatrix} \psi_2 & \varpi & \dots & \varpi \\ \psi_1 & \varpi & \dots & \varpi \\ 1 & \varpi & \dots & \varpi \\ 0 & \varpi & \dots & \varpi \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \varpi & \dots & \varpi \end{bmatrix}}_{M_2} \begin{bmatrix} \varepsilon_{t-2} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \dots \\
&\dots + \underbrace{\begin{bmatrix} \psi_q & \varpi & \dots & \varpi \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \varpi & \dots & \varpi \\ 0 & \varpi & \dots & \varpi \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \varpi & \dots & \varpi \end{bmatrix}}_{M_q} \begin{bmatrix} \varepsilon_{t-q} \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & \varpi & \dots & \varpi \\ \psi_q & \varpi & \dots & \varpi \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \varpi & \dots & \varpi \\ 0 & \varpi & \dots & \varpi \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \varpi & \dots & \varpi \end{bmatrix}}_{M_{q+1}} \begin{bmatrix} \varepsilon_{t-q-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \dots \\
&\dots + \underbrace{\begin{bmatrix} 0 & \varpi & \dots & \varpi \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \varpi & \dots & \varpi \\ \psi_q & \varpi & \dots & \varpi \end{bmatrix}}_{M_{q+1}} \begin{bmatrix} \varepsilon_{t-l-q} \\ 0 \\ \vdots \\ 0 \end{bmatrix}
\end{aligned}$$

where again ϖ denotes values of no interest for the analysis and $M_i \in \mathbb{R}^{l \times l}$. To find the recursive formula, use the knowledge that this MA process must be the generalized MA representation of the AR process we are interested in. Since for the AR process we are looking for $F\varepsilon_{t-1} = M_1\varepsilon_{t-1}$ it is obvious that $\psi_1 = \phi_1$. The first column of M_2 has been computed as the first column of F^2 in (5). Note that one can also think about the computation of the first column of M_2 as computing the first column of FM_1 . This means that one can compute the first column of M_3 by using the first column of FM_2 , which has the structure:

$$F^3 = \begin{bmatrix} \phi_1 & \dots & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \psi_2 & \varpi & \dots & \varpi \\ \psi_1 & \varpi & \dots & \varpi \\ 1 & \varpi & \dots & \varpi \\ 0 & \varpi & \dots & \varpi \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \varpi & \dots & \varpi \end{bmatrix}$$

thus, $\psi_3 = \phi_1\psi_2 + \phi_2\psi_1 + \phi_3$. Continuing with this approach, one finds

$$\begin{aligned}
\psi_4 &= \phi_1\psi_3 + \phi_2\psi_2 + \phi_3\psi_1 + \phi_4 \\
\psi_5 &= \phi_1\psi_4 + \phi_2\psi_3 + \phi_3\psi_2 + \phi_4\psi_1 + \phi_5 \\
&\vdots
\end{aligned}$$

As soon as ψ_q is reached, the structure is:

$$\begin{aligned}
0 &= \phi_1\psi_q + \dots + \phi_q\psi_1 + \phi_{q+1} \\
0 &= \phi_2\psi_q + \dots + \phi_{q+1}\psi_1 + \phi_{q+2} \\
&\vdots
\end{aligned}$$

This recursive relation resembles to one derived by using the lag-operator and assuming invertibility / $\tau = \infty$.

3 Improving Durbin's Method

I just demonstrated that for $j \geq q$, the j th parameter of the AR representation of an MA(q) process is generated by a q th order difference equation:

$$0 = \phi_{j-q}\psi_q + \cdots + \phi_{j-1}\psi_1 + \phi_j \Leftrightarrow \phi_j = - \sum_{i=1}^q \psi_i \phi_{j-i}$$

As noted in the introduction, the method of Durbin (1959) can be viewed as estimating this equation, viz estimating an AR on the parameters of the initial AR estimate. Note that the recursive formula implies that if $j \leq q$, the j^{th} parameter of the initial AR is not generated by the full difference equation. Therefore, using up to the q^{th} parameter of the initial AR for the second estimation will usually not improve the quality of the estimate since these observations are not generated by the process one seeks to estimate. The need to drop those estimates makes the usage of the information that $\phi_1 = \psi_1$ particularly helpful because the first parameters are often the ones estimated with the greatest accuracy.

A classic approach to include stochastic a priori information into OLS estimation is the f -class estimator provided by Theil (1963). Note the a priori information as

$$\underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}}_{\mathbf{R}} \psi = \hat{\phi}_1 + \epsilon$$

Where ϵ denotes the estimation error from using $\hat{\phi}_1$ instead of the true parameter. Denoting the variance of ϵ as $\sigma^2(\hat{\phi}_i)$, the f -class estimator for ψ is:

$$\hat{\psi}^r = - \left(\frac{1}{\sigma^2} \mathbf{X}'\mathbf{X} + \mathbf{R}' \frac{1}{\sigma^2(\hat{\phi}_i)} \mathbf{R} \right)^{-1} \left(\frac{1}{\sigma^2} \mathbf{X}'\mathbf{y} + \mathbf{R}' \frac{1}{\sigma^2(\hat{\phi}_i)} (-\hat{\phi}_1) \right)$$

where \mathbf{X} denotes matrix of explanatory variables, \mathbf{y} is the vector containing the depended variables and σ^2 is the variance of the disturbances. As usual, one has to replace σ^2 and $\sigma^2(\hat{\phi}_i)$ with their OLS estimates $\hat{\sigma}^2$ and $\hat{\sigma}^2(\hat{\phi}_i)$. The asymptotic variance of the estimates is given by the diagonal elements of

$$\left(\frac{1}{\sigma^2} \mathbf{X}'\mathbf{X} + \mathbf{R}' \frac{1}{\sigma^2(\hat{\phi}_i)} \mathbf{R} \right)^{-1}$$

Note that this gives a second reason for the exclusion of $\hat{\phi}_1$ from \mathbf{X} and \mathbf{y} : it ensures that the assumption of no correlation between ϵ and the innovations of the core regression model, which underlies the f -class estimator, is not violated.

4 Some remarks and a simulation

The recursive formula is helpful to think about the problems concerning MA processes with roots in the unit circle. It implies that such a process has an AR

representation whose parameters usually explode for $p \rightarrow \infty$. At the same time there exists a "sibling" MA process with the same first and second moment and all roots outside of the unit circle, implying a non explosive AR representation. Requiring $w_t = 0$ makes sure that the first and second moment are stationary, which tends to make an estimator pick the representation with non explosive parameters and thus the invertible MA process. This means that there is no hope in estimating the parameters of a noninvertible MA process with the help of the recursive formula by using some kind of restricted estimator. Since the parametrization one tries to estimate is unlikely, the estimator will use the remaining degrees of freedom to yield an MA process that is "as invertible as possible" given the restrictions and thus biased estimates.

That said, one should also appreciate the relative scale of the problem. Consider an MA process with one root that is just a little smaller than one, say 0.98. If a practitioner knows this from theory and is confronted with only a few hundred observations, he or she should not spend sleepless nights worrying about the bias induced by the noninvertibility. The reason is simple: the invertible sibling of the MA has the same roots except one that is $1/0.98$ instead of 0.98. Therefore its parameters are often very similar. In fact, estimates for an MA process with roots outside the unit circle except one might be as precise than the estimates for a process with the smallest root of, say, 1.001. The reason is as follows: The AR representation of an invertible MA process is de facto finite if one takes the precision of a digital computer as zero. However, such an AR representation of an MA process with a root of approximately one is very very large. This means that faced with a few hundred observations generated by a close-to unit root MA process, one can not avoid to specify an insufficient number of AR lags which distorts the parameter estimates. A reasonable AR representation of the invertible counterpart of an MA process with 0.98 as smallest root tends to be smaller and thus induces a smaller truncation bias. Therefore, the effects of the smaller truncation bias can counterbalance the general bias introduced by estimating the "wrong" process if a noninvertible MA process has only one root inside but close to the unit circle.

To illustrate this point as well as the gains of the using a restricted OLS estimator instead of Durbin's original estimator, I simulated both estimation procedures in the parameter space of an MA(2) process. I investigate the region $\psi_1, \psi_2 \in [-2.2, +2.2]$ which is divided into a grid with 221 points and partition the simulations for invertible and noninvertible processes. To study the gain when estimating invertible processes, all points in the grid that yield noninvertible processes are skipped. To study the performance when estimating noninvertible processes, points where either no root is in the unit circle or where the norm of the smallest root falls short of 0.8 are skipped.

For each point in the grid that is not skipped, 400 artificial observations with standard normal innovations are generated. An AR(100) is estimated using OLS and these estimates are used to estimate an AR(2) process with the Yule-Walker estimator and without dropping any estimate (Durbin's Method) and with Theil's f -class estimator using the restriction $\hat{\psi}_1 = \hat{\phi}_1$ while dropping $\hat{\phi}_1$ and $\hat{\phi}_2$ from the sample. The squared difference between the true parameters and estimates are computed and saved, after repeating this procedure 500

times for the particular point in the grid the mean of all squared differences is taken as estimate of the MSE. Figure 1 exhibits the results.

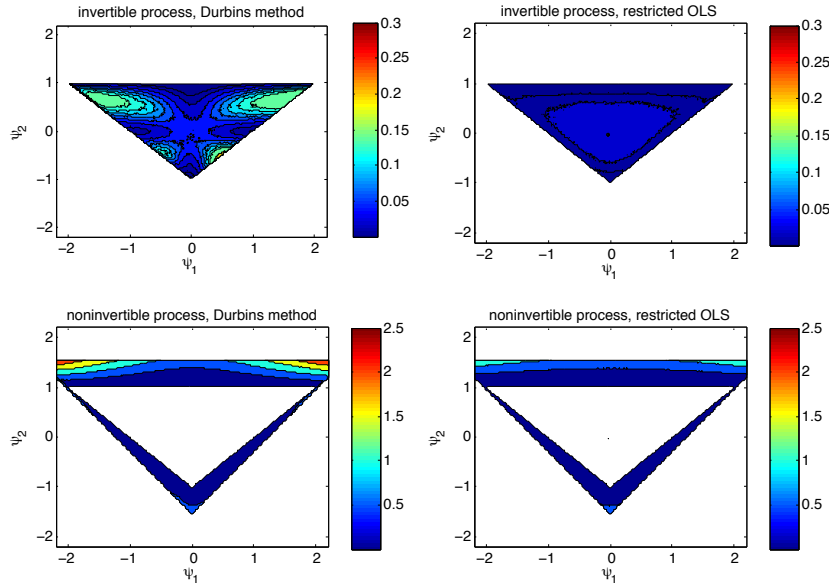


Figure 1: Simulation for the parameter space of an MA(2) process. The lower panel exhibits non invertible processes with roots between 0.8 and 1.

Note that in figure 1, there are small regions where Durbin’s methods yields slightly better results than restricted OLS. However, for wide areas of the parameter space restricted OLS yields a far smaller MSE. Further note that, as stated above, the MSE tends to increase quite moderate outside of the invertibility triangle.

A question of great practical importance is of what order the initial AR estimation should be. This question is tedious since it is a trade off: on the one hand, a larger order of the AR allows the sample which is used to estimate the MA process to be larger and tends to reduce the truncation bias. On the other hand, the precision of the estimates is the worse the more parameters have to be estimated. There is some fairly elaborate work on the question of the optimal trade-off by Broersen (2000). However, in most cases a simple rule of thumb works satisfactory well: make the AR process as large as possible but avoid less than 4 data points per estimate.

5 Generalization for Multivariate Processes

Denoting the parameters of a k -dimensional MA process as Ψ_j and the parameters of its VAR representation as Φ_i , it is easy to see that exactly the same recursive formulae apply for multivariate processes, given the assumption applies for each of the time series and the constants are zero. In this case one can express Φ_m with $m > q$ as:

$$\Phi_m = - \sum_{j=1}^q \Phi_{m-j} \Psi_j$$

Which means that, given $T \gg q$ estimates for the parameters of the VAR, the data model for the estimation of a multivariate MA process is

$$\underbrace{\begin{bmatrix} \hat{\Phi}_T \\ \hat{\Phi}_{T-1} \\ \vdots \\ \hat{\Phi}_{2q+1} \end{bmatrix}}_{\tilde{\Phi}} = - \underbrace{\begin{bmatrix} \hat{\Phi}_{T-1} & \hat{\Phi}_{T-2} & \dots & \hat{\Phi}_{T-q} \\ \hat{\Phi}_{T-2} & \hat{\Phi}_{T-3} & \dots & \hat{\Phi}_{T-q-1} \\ \vdots & \vdots & & \vdots \\ \hat{\Phi}_{2q} & \hat{\Phi}_{2q-1} & \dots & \hat{\Phi}_{q+1} \end{bmatrix}}_{\Phi \in \mathbb{R}^{k(T-2q-1) \times kq}} \underbrace{\begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_q \end{bmatrix}}_{\tilde{\Psi}} + u$$

It is easy to see that one can divide this into k single standard regression models of the kind $\phi_j = -\Phi\psi_j + u_j$, where ϕ_j is the j th column of $\tilde{\Phi}$ and ψ_j is the j th column of $\tilde{\Psi}$, implying that the unrestricted OLS estimate is

$$\hat{\psi}_j = -(\Phi'\Phi)^{-1}\Phi'\phi_j$$

However, for multivariate processes it is also very useful to exploit the knowledge that $\Psi_1 = \hat{\Phi}_1 + \epsilon$. Since we estimate k single regression models, one has to use a different restriction for each estimation. For estimation j note the a priori information as

$$\underbrace{\begin{bmatrix} I_k & O \end{bmatrix}}_{\mathbf{R}} \psi_j = \underbrace{\begin{bmatrix} I_k & O \end{bmatrix}}_{\mathbf{r}_j} \phi_j + \epsilon_j$$

where O is a matrix of zeros in $\mathbb{R}^{k \times k \cdot (p-1)}$. This allows to note the f -class estimator for estimation j as

$$\hat{\psi}_j^r = - \left(\frac{1}{\sigma^2} \Phi'\Phi + R'\Sigma(\Phi_1)_j^{-1}R \right)^{-1} \left(\frac{1}{\sigma^2} \Phi'\phi_j + R'\Sigma(\Phi_1)_j^{-1}(-\mathbf{r}_j) \right)$$

Where $\Sigma(\Phi_1)_j$ is a matrix of zeros with the variances of the OLS parameter estimates for the j th column of Φ_1 on its main diagonal. There is no covariance as long as one uses the standard OLS estimator for VARs, as described e.g. in Lütkepohl (2005, pp 71). This estimator can be obtained by rewriting the VAR and transforming it into a standard linear regression model with the help of the Vec operator. As pointed out by Lütkepohl, this is identical to an OLS estimation of the k equations separately. This means that the estimation does not provide information about the covariance between the elements in the columns of Φ_i , we therefore only use the variances. Similar to the univariate case, the variances of the MA estimates can be computed using just the left part of the estimator.

6 Conclusion

This paper provides a simple method to estimate both univariate and multivariate MA processes by exploiting the recursive relation between the MA process and its AR representation. A simulation study for the parameter space of an MA(2) process indicates that the method tends to have a smaller MSE than Durbin's method and is relatively robust with respect to unit roots of the MA process, in the sense that the MSE increases moderately outside the invertibility triangle.

References

- P.M.T Broersen. Autoregressive model orders for Durbin's MA and ARMA estimators. *Signal Processing, IEEE Transactions on*, 48(8):2454–2457, 2000.
- J Durbin. Efficient Estimation of Parameters in Moving-Average Models. *Biometrika*, 46(3/4):306–316, 1959.
- James Douglas Hamilton. *Time Series Analysis*. Princeton University Press, 1994.
- Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, 2005.
- Niclas Sandgren and Petre Stoica. On Moving Average Parameter Estimation. Technical Report 2006-22, Department of Information Technology, Uppsala University, 2006.
- Henri Theil. On the Use of Incomplete Prior Information in Regression Analysis. *Journal of the American Statistical Association*, 58(302):401–414, 1963.
- Dag Tjøstheim and Josten Paulsen. Bias of some Commonly-Used Time Series Estimates. *Biometrika*, 70(2):389–399, 1983.